



Tracking seasonal changes in diversity of pollen allergen exposure: Targeted metabarcoding of a subtropical aerobiome

B.C. Campbell^a, J. Al Kouba^b, V. Timbrell^a, M.J. Noor^c, K. Massel^d, E.K. Gilding^d, N. Angel^e, B. Kemish^e, P. Hugenoltz^e, I.D. Godwin^d, J.M. Davies^{a,f,*}

^a Queensland University of Technology, Brisbane, Australia

^b Macquarie University, Sydney, Australia

^c Fatema Jinnah Women University, Rawalpindi, Pakistan

^d The University of Queensland, Brisbane, Australia

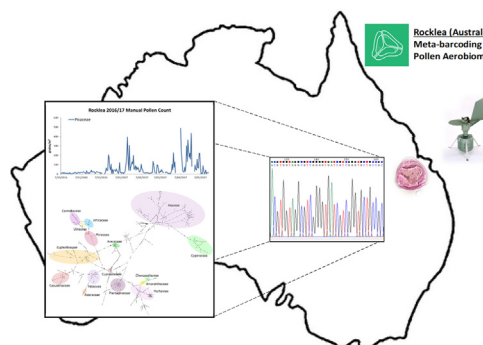
^e Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

^f Metro North Hospital and Health Service, Brisbane, Australia

HIGHLIGHTS

- Pollen is a primary cause of Allergic Rhinitis in Rocklea, Queensland, Australia.
- Exposure to subtropical pollen is difficult to discern by light microscopy.
- *rbcl* meta-barcoding of the pollen season identified several taxa to genus levels.
- Chloridoideae and Panicoideae pollen were the dominant Poaceae pollen source.
- Sequencing of routinely collected pollen may track seasonal changes in diversity.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 24 March 2020

Received in revised form 21 July 2020

Accepted 21 July 2020

Available online 25 July 2020

Editor: Pavlos Kassomenos

Keywords:

Aerobiome

Allergic rhinitis

Metabarcoding

Next generation sequencing

Pollen-monitoring

Subtropics

ABSTRACT

The importance of grass pollen to the global burden of allergic respiratory disease is well established but exposure to subtropical and temperate pollens is difficult to discern. Current monitoring of airborne pollen relies on light microscopy, limiting identification of taxa to family level. This informs seasonal fluctuations in pollen aerobiology but restricts analysis of aerobiological composition. We aimed to test the utility of DNA metabarcoding to identify specific taxa contributing to the aerobiome of environmental air samples, using routine pollen and spore monitoring equipment, as well as assess temporal variation of Poaceae pollen across an entire season. Airborne pollen concentrations were determined by light microscopy over two pollen seasons in the subtropical city of Brisbane (27°32'S, 153°00'E), Australia. Thirty daily pollen samples were subjected to high throughput sequencing of the plastid *rbcl* amplicon. Amplicons corresponded to plants observed in the local biogeographical region with up to 3238 different operational taxonomic units (OTU) detected. The aerobiome sequencing data frequently identified pollen to genus levels with significant quantitative differences in aerobiome diversity between the months and seasons detected. Moreover, multiple peaks of Chloridoideae and Panicoideae pollen were evident over the collection period confirming these grasses as the dominant Poaceae pollen source across the season. Targeted high throughput sequencing of routinely collected airborne pollen samples appears to offer utility to track temporal changes in the aerobiome and shifts in pollen exposure. Precise identification of the composition and temporal distributions of airborne pollen is important for tracking biodiversity and for management of allergic respiratory disease.

© 2020 Elsevier B.V. All rights reserved.

* Corresponding author at: Queensland University of Technology (QUT), Level 11 Bancroft Building, QIMR Berghofer 300 Herston Road, Herston, QLD 4101, Australia.
E-mail address: j36.davies@qut.edu.au (J.M. Davies).

1. Introduction

Allergic rhinitis (AR) afflicts 500 million people worldwide, adversely impacting quality of life (Katelaris et al., 2012; Pawankar et al., 2011). The importance of grass pollen to the global burden of AR has been well established. While subtropical and temperate grass pollen allergens show distinct immunological reactivity, the environmental exposure to different grass pollens remains unknown (Davies et al., 2005; Davies et al., 2011a; Nony et al., 2015; Weber, 2007; Kailaivasan et al., 2020). With climate change it is predicted that the burden of AR will increase and that subtropical climate zones will expand, thus increasing the contribution of subtropical grasses to allergic disease (Beggs and Bennett, 2011; Morgan et al., 2011). Using current pollen monitoring and forecasting methods, the biogeographical distribution of species, their composition across regions, and their seasonal fluctuations remain unknown. This is especially the case in subtropical regions where the identification of new and novel allergens is a recent and ongoing process (Campbell et al., 2015; Davies et al., 2008; Davies et al., 2011a; Davies et al., 2011b). More precise identification of local pollen exposures that contribute to allergic respiratory disease burden would be advantageous for better self- and clinical management of AR patients.

Currently, pollen monitoring methods depend on light microscope-based identification and counting of pollen in collected air samples. This is highly labour-intensive and for most types of pollen, including grass, is not effective in classifying pollen beyond family level (Hirst, 1952; Thibaudon et al., 2013). The inability to identify pollen beyond family level reduces the accuracy of the existing pollen monitoring process, whereby the pollen of numerous species, particularly the grasses, is morphologically identical by light microscopy. Accurate identification of grass pollen to subfamily or further to the genus level is important because the pollen of separate subfamilies differ in the specificity of the immune response they elicit for patients depending on their sensitisation profile (Chabre et al., 2010; Davies et al., 2011a; De Weger et al., 2011; Etto et al., 2012; Nony et al., 2015; Kailaivasan et al., 2020).

A range of alternative techniques for pollen classification have been pursued including digital image based-detectors (Landsmeer et al., 2009), spectroscopic techniques (Rittenour et al., 2012) and DNA-based approaches (Longhi et al., 2009).

With the advent of Next Generation Sequencing (NGS) combined with DNA metabarcoding (Bell et al., 2016b; Cristescu, 2014), DNA-based analysis of aerobiological surveys has shown great promise as a technique for identification and quantification of mixed pollen and environmental samples (Keller et al., 2015; Kraaijeveld et al., 2015; Núñez et al., 2017; Leontidou et al., 2018; Brennan et al., 2019). Metabarcoding involves analysis of amplified fragments of DNA (barcode) from well conserved genes in a heterogeneous biological sample. These barcodes should have highly similar sequences within taxa and yet contain enough variability to allow taxonomic identification. The capacity for pollen DNA metabarcoding to be transferred from laboratory experiments or small-scale environmental trials, to its application across a complete pollen season is only now being tested.

Kraaijeveld et al. (2015) demonstrated that metabarcoding in combination with NGS provided an accurate qualitative and quantitative view of the species composition applied to a small number of environmental airborne pollen samples. This method identified individual grass genera present in a mixed sample of grass pollen. With regards to routine pollen monitoring apparatus, Núñez et al. (2017) established that high-throughput DNA sequencing applied to aerobiological samples obtained with Hirst spore traps, can provide reliable results and be implemented to monitor a diverse array of prokaryotic and eukaryotic species.

While several studies report characterising pollen sources across a limited sample set, few studies have applied metabarcoding across an

entire pollen season nor examined the seasonality of pollen exposure using routine environmental air samples collected with a Hirst-style pollen and spore trap. Brennan et al. (2019) undertook a well-defined study, using a cyclonic sampler, of the temperate grass pollen season of Great Britain with the intent to explore whether temporal turnover in species composition match terrestrial flowering or whether species richness steadily accumulates over the grass pollen season. They discovered that grass species exhibited distinct, temporally constrained peaks of pollen release, which varied with latitude and longitude throughout Great Britain, showing that the taxonomic composition of temperate grass pollen exposure changes across the grass pollen season. We have studied a subtropical region for which there are no data on airborne species composition. The present study aimed to test the applicability of an emergent metabarcode technique for the analysis of the pollen aerobiome from routinely collected environmental air samples over a complete pollen season using a Hirst-style continuous flow impactation sampler, with the goal of classifying taxa to the genus level. A further aim was to examine whether this method could explore seasonal variation in an exemplary subtropical region to both track the grass pollen season, compare the accuracy of manual pollen counting versus metabarcoding as well as reveal the broad genus level diversity of a little researched subtropical environment. Specifically, we examine whether multiple peaks in airborne grass pollen observed (Green et al., 2002; Medek et al., 2016) were due to flowering of multiple genera successively or repeated flowering of the same Poaceae genera over the season.

2. Material and methods

2.1. Sampling site

Aerobiological samples were collected from a subtropical region in south-east Queensland, Australia using a continuous flow volumetric Burkard seven day pollen and spore trap (Burkard Manufacturing, UK) located within a peri-urban suburb of the city of Brisbane called Rocklea (latitude (S) 27°32'; longitude (E) 153°00'). This area has an elevation of 13 m above sea level and the monitor height was placed 4 m above the ground. According to Green et al. (2002), the vegetation around the Rocklea site is typical of that of suburban Brisbane and is composed of endemic and exotic species. The sampling site is located adjacent to a grassy paddock surrounded by an open eucalypt woodland as well as herbaceous and riparian communities.

Pollen was monitored daily over a 24-h period and collected at 9 am each morning. There were two data collection periods for this study consisting of a 12-week pilot project that took place during the summer and autumn of the 2015 pollen season. Twelve daily samples were analysed. The monitoring period of the 2016/17 pollen season encompassed 225 days between 07.10.2016 to 18.05.2017. To represent the 2016/17 pollen season, one daily sample every seven days, based on consistent sample availability, was analysed.

2.2. Pollen collection

Pollen was collected with a 7-day drum on the lid assembly of the volumetric pollen and spore trap (Burkard Manufacturing, United Kingdom) used with Melinex® tape coated with a silicon-based adhesive (Lanzoni s.r.l., Bologna, Italy). The exposed Melinex® tape was cut into a 24-h segment. After 24 h of monitoring, the tape was collected and half was cut longitudinally and placed into a 2 mL screw cap tube and stored at −80 °C for future DNA extraction, the other half of the tape was mounted on glass slides with fuchsine stained gelatine mountant (Lanzoni s.r.l., Bologna, Italy) in preparation for manual counting using light microscopy. Collected airborne pollen impacted onto the tape was counted over four longitudinal transects at 400× magnification. Pollen and spore counts were then converted to grains/m³ of air according to the following calculation:

$$[\text{Pollen or Spore}] = n \left[\frac{L \times W}{L \times \left(\frac{F}{M} \right) \times N} \times \frac{1}{D \div 1000 \times t} \right]$$

where; [Pollen or Spore] - Atmospheric pollen grain or fungal spore concentration per cubic metre, *n* - Number of pollen grains or fungal spores counted in the analysed area of the microscope slide, *L* - Length of impaction area on the melinex tape or glass slide (mm) e.g., 48 mm, *W* - Width of impaction area on the melinex tape or glass slide (mm) e.g., 14 mm, *F* - Field number written on the eyepiece (typically a number between 18 mm and 25 mm), *M* - Objective magnification (typically 40), *N* - Number of transects e.g. 4, *D* - Flow rate of the equipment (L/min) e.g., 10L/min.t - Duration of the sampling period (min) e.g., 1440 min, and expressed as a daily mean concentration (Beggs et al., 2018).

2.3. Pollen identification (light microscopy)

Pollen of twelve families (Fig. 1) based on abundance or allergenic interest were identified and counted according to The Australasian Pollen and Spore Atlas (APSA), <http://apsa.anu.edu.au>.

2.4. DNA extraction

DNA extraction was based on a modified hexa decyltrimethylammonium bromide (CTAB) technique, as described by Rogers and Bendich (1985). Each sample consisted of a 2 mL screw cap tube containing half of the aerobiological sample. The Melinex® tape was carefully removed from the tube with sterile forceps and cut using sterile scissors longitudinally and then further into six equal pieces (12 total), ensuring each piece was placed back into the original tube. To each tube 500 µL of CTAB buffer (2% CTAB, 100 mM Tris-HCL pH 8.0, 20 mM EDTA, 1.4 M NaCl, 3% polyvinylpyrrolidone, 0.2% β-mercaptoethanol) and 0.2 g of glass beads (425–600 µm diameter) was added. Each tube was then placed into a Mini-Beadbeater-16 (BioSpec Products, Inc. Oklahoma) and the samples were lysed using two 1-min cycles (3450 oscillations/min). Once homogenisation was complete, samples were incubated at 68 °C for an hour and then centrifuged for 20 min at 15,120g (13,000 rpm). The supernatant was transferred to new tubes whereby a half volume of chloroform:isoamyl alcohol (24:1) was added per sample and then spun for a further 20 min. The aqueous phase was transferred to new tubes and DNA was precipitated with 1 volume of cold absolute ethanol and 1/10th volume of 7.5 M ammonium acetate was added per sample. Samples were then allowed to incubate overnight in a – 20 °C freezer to allow for

further DNA precipitation. The following day, DNA was pelleted by centrifugation for 20 min at 13,000 rpm. All subsequent steps describing DNA pellet drying and resuspension followed the protocol of Rogers and Bendich (1985) and samples were resuspended in 50 µL of DNase/RNase free H₂O.

2.5. PCR amplification

The large subunit of RuBisCo (*rbcL*) in plant chloroplasts was targeted using modified primers designed to contain Nextera adapter sequences (*rbcL*_1F: 5'- **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**ATGTGTCACCACAAACAGAGACTAAAGC-3' and *rbcL*_599R: 5'-**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GTAATCAAGTCCACCRG-3'), where the Nextera adaptor sequence is bolded and underlined (CBOL Plant Working Group A DNA barcode for land plants, 2009). *rbcL* was a barcode recommended to reach the genus level of classification (Kress et al., 2005; CBOL Plant Working Group A DNA barcode for land plants, 2009). Because *rbcL* is a plastid marker, we hypothesised it would be well detected within a pollen grain compared with any single gene in the haploid nuclear genome Tang et al. (2009). Based on our *in silico* analysis of plants observed in the Rocklea region by Green et al. (2002) we observed that all species present in the local environment could be distinguished to genus using the *rbcL* amplicon.

The length of target sequence varied depending on the species from approximately 550–590 bp. The *rbcL* PCR was performed in a 20 µL reaction mixture containing 2 µL of extracted DNA template, 1 × 10× Pfx amplification buffer, 1/10 volume of 10× Pfx Enhancer solution, 1.2 mM MgCl₂, 0.3 µM each primer, 0.3 mM dNTPs and 1 U Platinum® Pfx DNA Polymerase (ThermoFisher Scientific). The PCR reaction program consisted of: one cycle at 94 °C, 5 min; 40 cycles at 94 °C, 20 s; 55 °C, 20 s and 68 °C, 45 s; one cycle at 68 °C, 2 min; and held at 4 °C. All PCR products were then purified utilising the Wizard® SV Gel and PCR Clean-Up System (Promega) according to manufacturer's instructions, in preparation for Illumina sequencing.

2.6. Illumina library preparation and sequencing

Preparation of the Illumina library was performed using the workflow outlined by Illumina (#15044223 Rev.B). Purified DNA was then indexed with unique 8 bp barcodes using the Illumina Nextera XT 384 sample Index Kit A-D (Illumina FC-131-1002) in standard PCR conditions with NEBNext® Ultra™ II Q5® Mastermix. Indexed amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter) and pooled in equimolar concentrations and sequenced on

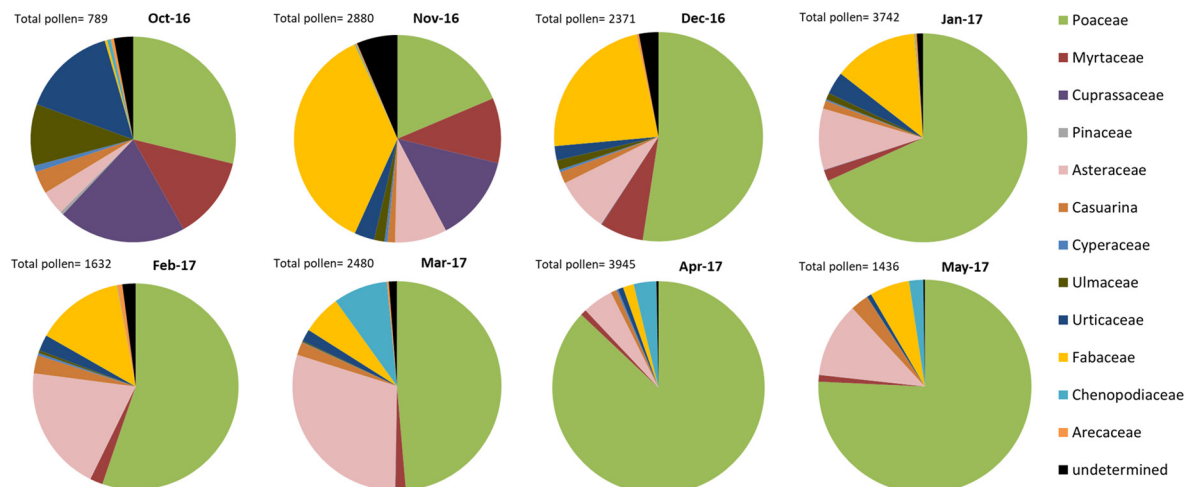


Fig. 1. Pie charts showing the distribution of manual pollen counts according to family-level classification per month across the Rocklea region 2016/17.

the MiSeq Sequencing System (Illumina) using paired end sequencing with V3 2 × 300 bp chemistry at the Australian Centre for Ecogenomics according to manufacturer's protocol.

Control reactions incorporated in the amplicon library construction and sequencing included: i) Positive amplification control from a known mock community to monitor for bias in the amplicon library construction, ii) Negative amplification control from a like processed reagent control to monitor for contamination in library construction, iii) Single well empty chamber controls within processing plates to monitor for cross contamination within the library preparation and iv) Negative control of four random well locations for sets of index positions between runs to monitor for run to run bleed through contamination within the instrumentation flow lines, designated as in-line controls.

Samples producing a minimum of 10,000 raw reads with an average Q30 of >70% over 600 bp paired end reads were progressed to data analysis.

All sequencing data has been uploaded to the National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/> with the corresponding BioProject ID number PRJNA579003.

2.7. Data analysis

Prior to data analysis, a custom *rbcl* reference library including 51 species observed in the Rocklea region was compiled (Green et al., 2002; Moss, 1967).

Analysis using an Agilent TapeStation of amplicons prior to sequencing, indicated that sequences of 590 bp including the primer sequences could be expected. This meant that the entire length of both forward and reverse reads were required, including the end regions which showed higher variation in sequence quality than ordinarily accepted. To minimize errors due to low quality read ends prior to merging, the paired end sequences were trimmed using Trimmomatic (Bolger et al., 2014) applying a quality filter whereby a sliding window of 4 bases was checked for an average quality score of ≥ 20. Reads which failed this criterion or whose length were less than 50% of the expected length were discarded. Sequences were merged using PANDAseq (Masella et al., 2012) and the primer sequences were trimmed off.

A standard Open Reference OTU picking (pick_open_reference_otus.py; Rideout et al., 2014) from QIIME (Caporaso et al., 2010) was used, with 99% similarity against our custom *rbcl* reference library (Software parameters; Table S1, Supporting information).

OTUs with a minimum read count of 20 or less were excluded. Reads were aligned using BLASTn against the NCBI nucleotide database (Camacho et al., 2009; accessed on the 2nd of June 2018) utilising default parameters. The results were curated to select the best hits based on the highest bit score in conjunction with having an identity ≥97% and 100% coverage. Taxa with the same genus identification were combined, while those that had nucleotide alignments showing equal percentage identity with different plant classifications were further scrutinised based on records in the database Atlas of Living Australia (ALA; www.ala.org.au). Those taxa with records at or near the Rocklea site were chosen over other possible genera for classification. At least 27% of total OTUs studied required curation against ALA records. Sequences for remaining taxonomic units that could not be confidently assigned to a genus were classified to the family level. Calculations of temporal changes in pollen diversity across the months and seasons were conducted utilising the R-package 'vegan' (Oksanen et al., 2013) which provides tools for diversity and dissimilarity analysis and ordination for community ecology, namely principal component (PCA) and redundancy analysis (RDA). Specifically, the PERMANOVA package was utilised to perform multivariate and two-way ANOVAs using distance matrices, of the aerobiological OTU data with method set to Euclidean (method = 'euc') (Anderson, 2005). Likewise, calculations of indicator values (fidelity and relative abundance) according to season and month were conducted utilising the indval function in the R-

package 'labdsv' (Roberts, 2012). First, an unconstrained PCA was performed to determine the largest gradient of variability across the samples through the calculation of synthetic variables (principal components). This was followed by performing a constrained RDA ordination plot to assess if the explanatory variables, namely month or season, can explain the variation across the dataset. Sequence alignment and editing was performed in Geneious Pro v. 7.1.9 (<https://www.geneious.com>) using the Muscle Alignment option with default parameters. Phylogenetic tree construction was performed using FastTree 2.1 (Price et al., 2010) and PopART (Population Analysis with Reticulate Trees) was utilised to construct median-joining networks (Bandelt et al., 1999; Leigh and Bryant, 2015).

3. Results

3.1. Pilot study outcomes of 2015–2016 pollen season

In total 899,237 *rbcl* reads were obtained in the pilot study, which accounted for an average of 74,936 [standard deviation (SD): 47,367; median: 85,586] filtered reads per sample over the 12 environmental samples. The 32 pollen taxa detected in 13 samples (12 daily + 1 control) mostly corresponded to plants observed in the local region. Amplified products could be classified to the genus level (Fig. S1B, Supporting information). Comparisons between pollen concentration data and metabarcoding indicated that DNA read counts emulated the family distributions for Poaceae, Asteraceae and Casuarinaceae, but there were some observable differences (Fig. S1A, Supporting information). Metabarcoding of routinely collected environmental pollen samples allowed the detection of several other plant families that were not identified by light microscopy. Of particular note were the Fabaceae and Euphorbiaceae families represented by *Albizia lebeck*, and *Acalypha* spp. respectively, the latter of which is not a family counted by microscopy. The outcomes of the pilot suggest that the methodology could be more broadly applied to environmental air samples collected across a pollen season with Burkard pollen and spore sampler.

3.2. Sequencing output across the 2016–2017 pollen season

Over the 2016/17 pollen season 299,544 reads were obtained over the 30 samples with an average of 9984 [standard deviation (SD): 4911; median: 9612] filtered reads per sample. The total number of individual OTUs identified after filtering was 3059 for the 2016/17 pollen season, with a sequencing read count per OTU ranging from 96,234 to 2 reads. For testing the utility of the technique for detailed analysis of pollen diversity and temporal shifts throughout the pollen season, we chose a count threshold of greater than 20 OTU reads. This resulted in a sub-sample population of 345 OTUs being further scrutinised.

3.3. Summary of the 2016/17 pollen season (light microscopy)

Across the 225 days of the Rocklea pollen season between October (2016) and May (2017), we categorised pollen as belonging to 12 different botanical families by light microscopy (Fig. 1). Pollen not belonging to these families were recorded as undetermined pollen. Total pollen concentrations per month varied from 789 grains/m³ for October (2016) through to 3945 grains/m³ in April (2017). Apart from the month of November (2016) that had a substantial count of Fabaceae present, Poaceae pollen predominated throughout the season with 1564 grains/m³ and comprised 55% of all taxa enumerated (Fig. 4A). In particular, from December (2016) through to May (2017), grass species made upwards of half of all species counted, with the months of April and May (2017) consisting of more than 75% Poaceae (Fig. 1). In descending order of pollen concentration throughout the season, Fabaceae (377 grains/m³; 13.28%), Asteraceae (342 grains/m³; 12%), Myrtaceae (102 grains/m³; 3.6%) and Cupressaceae (101 grains/m³; 3.6%) were the other significant families counted in Rocklea.

The pollen season for Poaceae displayed multiple peaks of abundance of airborne grass pollen with the highest peaks observed in mid-January (Summer; 2017) and early to mid-April (Autumn; 2017) (Fig. 2). The other families showed a distinct pollen season peak with Fabaceae, Myrtaceae and Cupressaceae flowering in spring around early to mid-November (2016), and Asteraceae flowering in autumn around early to mid-March (2017) (Fig. 2).

3.4. Analysis of the composition of routinely collected aerobiological samples

The majority of OTUs from the 2016/17 Rocklea aerobiome could be classified to the genus level (Fig. 3). Compared to light microscopy, metabarcoding detected the twelve families counted by microscopy as well as 34 additional families (Fig. S2, Supporting Information).

Metabarcoding results mirrored the abundance of genera per family identified via microscopy in several cases, with Poaceae taxa being the most frequently counted and comprising 38% of all pollen classified when being compared to the 12 families recorded by microscopic counting (Figs. 1, 3 and 4A). However, taxa classified as Fabaceae and Asteraceae by light microscopy, contributed a more substantial proportion of the pollen count than detected with metabarcoding, 13.28% vs 4.60% and 12% vs 0.65% respectively. Such differences in proportions may be explained by the presence of pollen belonging to other families not identified for which pollen was assigned as “undetermined” by light microscopy, such as Euphorbiaceae, which was the second most

frequent family detected via metabarcoding, followed by others such as Plantaginaceae. In general, the difference in proportions between the two techniques may be partly explained by classification of OTUs into the 34 other families (Fig. 4B).

However, there was a low proportion of Asteraceae identified with metabarcoding in the 2016/2017 season. While Asteraceae was the third most prominent family identified by microscopy, it scored a relatively low read count of 42 with sequencing (Fig. S2, Supporting information).

3.5. Tracking temporal changes in pollen diversity

When considered at the family level of classification, a relationship between read counts from the metabarcode data and pollen abundance for those taxa identified via light microscopy was evident for Poaceae ($r_s = 0.42$; p -value = 0.0224) and Myrtaceae ($r_s = 0.57$; p -value = 0.001), Cupressaceae ($r_s = 0.90$; p -value = 6.79e-12), Casuarinaceae ($r_s = 0.42$; p -value = 0.02), Urticaceae ($r_s = 0.38$; p -value = 0.04) but not for all of 12 families combined ($r_s = 0.23$; p -value = 0.45) (Fig. S3, Supporting information). The ability of metabarcoding to differentiate species distribution and aerobiological composition over time was evident by performing multivariate analysis of variance using distance matrices upon pollen OTUs.

Utilising month or season as treatment variables, we detected significant temporal changes ($P \leq 0.001$) (Table S2, Supporting information) in temporal distribution of airborne pollen composition. Principal

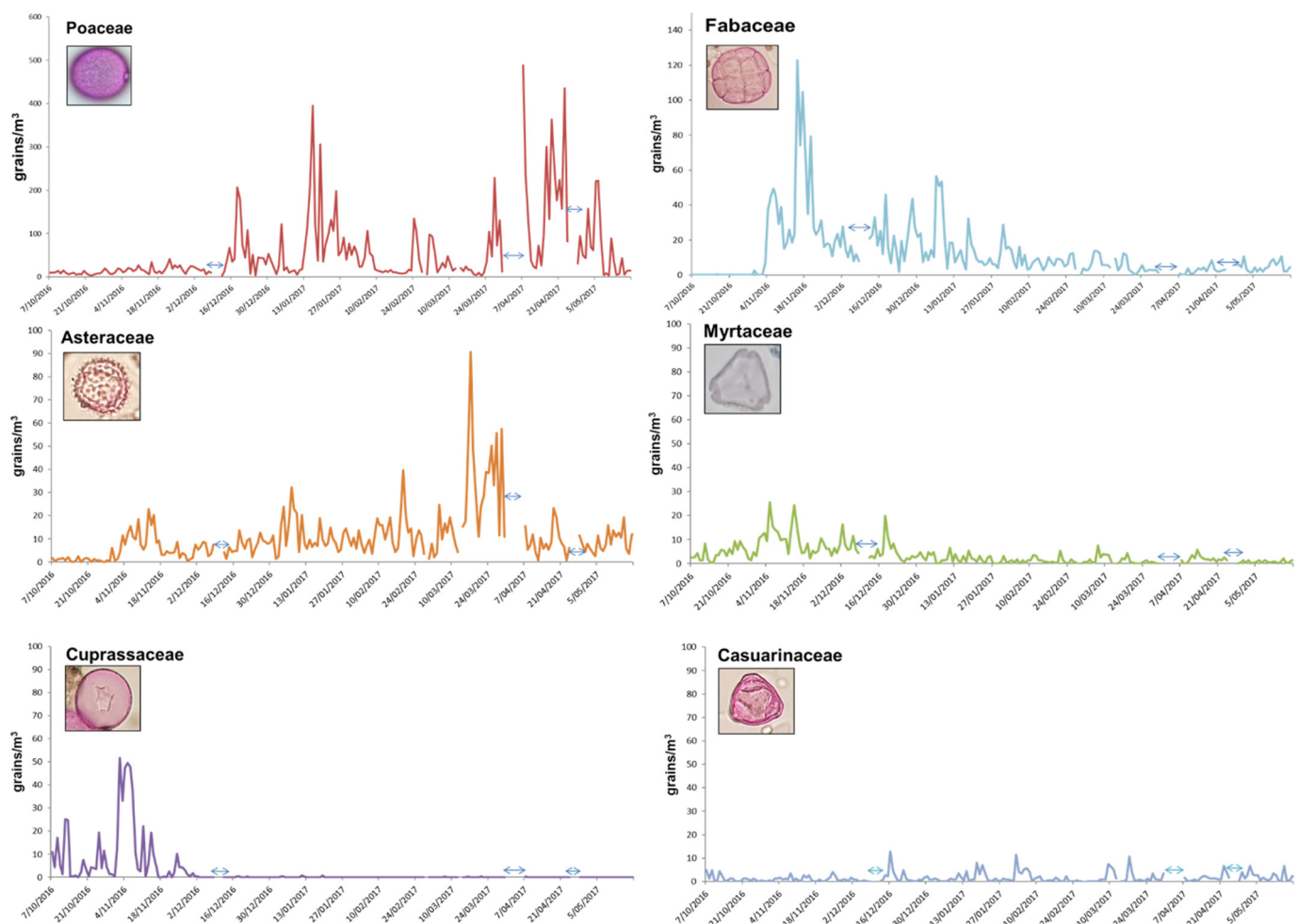


Fig. 2. Pollen Concentrations (grains/m³) counted by light microscopy for six of the most prominent families at Rocklea during the 2016/17 pollen season. Blue arrows indicate missing data.

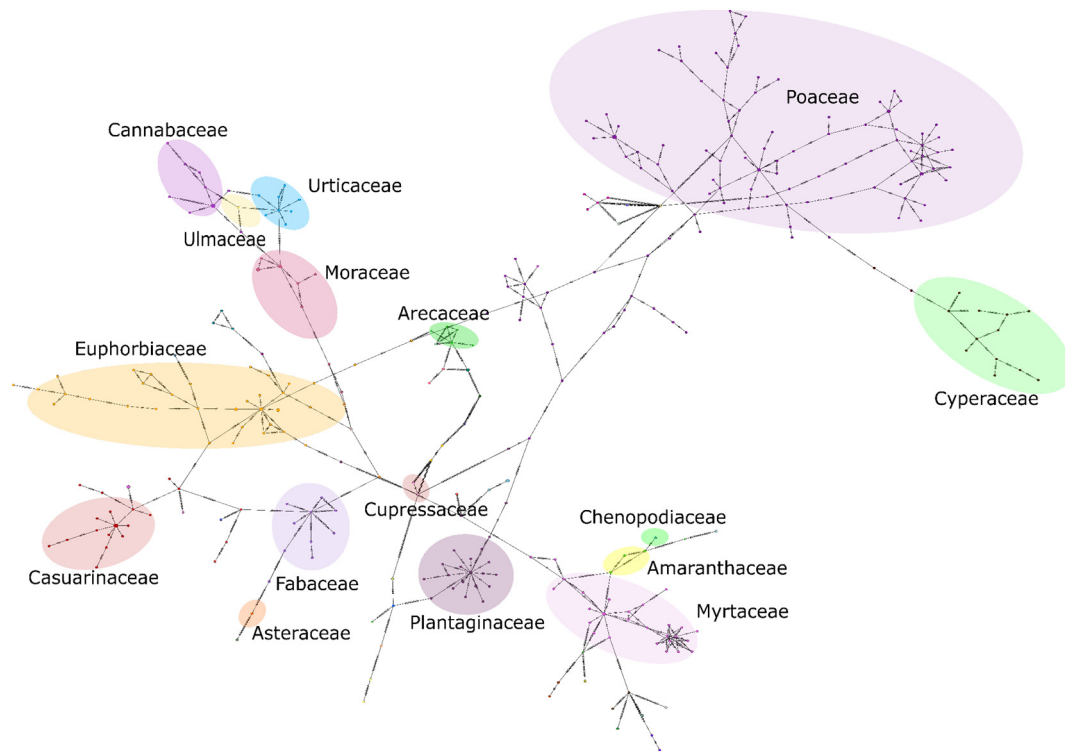


Fig. 3. Median-joining network of operational taxonomic units (OTU) present within the Rocklea 2016/17 pollen aerobiome clustered utilising the *rbcl* metabarcode. Each individual node represents an individual OTU and each different colour denotes a different family-group.

components analysis (PCA) showed that many of the samples clustered together based on their OTU compositions from the pollen samples clustered according to the months within which the pollen samples were collected, such as October, February, March, April and May (Fig. 5A&B). Sample clusters were also differentiated according to season, particularly for spring and autumn, with principal component 1 and 2 explaining 35.1% of the variation. However, comparing the R^2 values of the factors 'month' (0.38) or 'season' (0.18) to the R^2 values of the residuals, 0.62 and 0.82 respectively, shows that there was still high variability not explained by 'month' or 'season' (Table S2, Supporting information). Utilising both month and season as explanatory variables, multivariate analysis (RDA), explained 62.5% of the variation according to month and 99.4% of the variation according to season, forming unique clusters related to time of the season (Fig. 5A&B). While the RDA plots separated the points better than the PCA plot, and season better than month, there were observable overlaps. The temporal effects of month and season explained some of the variation for 'month' and a substantial level of the variation for 'season' on the ordination because the % RD-1 and RD-2 axes were high, which demonstrates the significant effect these constrained variables have, particularly for 'season', upon the composition and abundance of the pollen.

Analysis of indicator or discriminating (OTU) species, based on taxa whose relative abundance acts as a proxy for a particular season or month, revealed that this could only be undertaken accurately for specific seasons. Examples for the season of spring, showed both *Callitris* spp. (cypress; Cupressaceae) and *Typha* spp. (bulrush; Typhaceae) were strong indicator species for this season with indicator values of 0.875 ($P \leq 0.001$) and 0.7316 ($P \leq 0.001$), respectively (Table S3, Supporting information). This result was verified by heat map data of OTU abundance, with sequence reads peaking in the months of October and November (Fig. 4B). Likewise, for the season of autumn (March to May), Casuarinaceae spp. showed indicator species values as high as 0.6928 ($P \leq 0.001$) (Table S3, Supporting information & Fig. 4B). There were no indicator species that clearly defined the season of summer. The closest OTU was that of *Eleusine indica* (crowsfoot grass) with an

indicator value of 0.6001 ($P \leq 0.005$). But close inspection of the spread of sequence reads revealed there were low numbers present in the month of November 2016 (spring).

3.5.1. Utility of metabarcoding for tracking temporal changes in composition of grass pollen genera

Metabarcoding allowed a detailed breakdown of the classification and distribution of grass species within the Rocklea aerobiome. We observed that *Chloris gayana* (Rhodes grass; Sub-family: Chloridoideae) was the most prolific flowering species, not just within Poaceae but of all taxa analysed with 97,602 reads (Fig. 6A). Other prominent grasses were *Cynodon dactylon* (Bermuda grass; Sub-family: Chloridoideae) with 13,421 reads, *Megathyrus maximus* (Guinea/Green Panic grass; Sub-family: Panicoideae) with 10,843 reads and *Paspalum* spp. (commonly *P. notatum*; Sub-family: Panicoideae) with 5888 reads.

Median-joining network analysis showed that grass taxa clustered together according to sub-family, with Panicoideae having the most taxa identified but Chloridoideae had taxa with the majority of sequence reads (112,729 reads; Fig. 6B). Members from the Pooideae sub-family were observed less frequently, with the genus *Triticum* having the highest read count (2546).

When comparing peak flowering times observed by light microscopy of pollen with heat map data for the grass aerobiome, there is a corresponding pattern between the two datasets with peak flowering times in December (2016), January (2017), March (2017), April (2017) and May (2017) observed (Fig. 6A). The data show that several of the key flowering species, particularly *Chloris gayana* and *Cynodon dactylon*, have no discrete flowering time but repeatedly contribute to multiple pollen peaks during the summer and autumn seasons.

4. Discussion

This study has demonstrated that metabarcoding of routinely collected environmental airborne pollen samples has potential to monitor

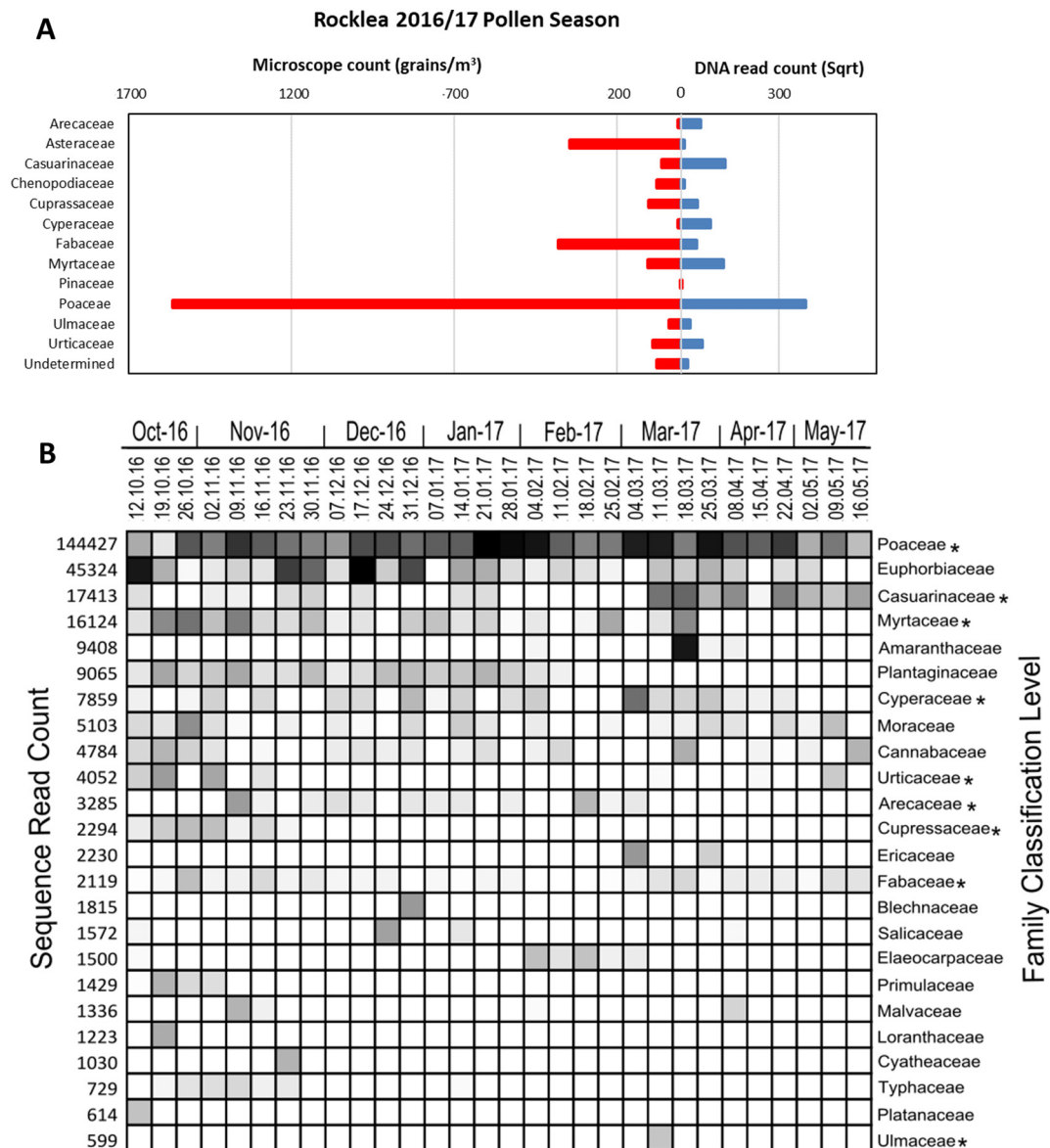


Fig. 4. 2016/17 Rocklea pollen season data. (A) Comparison of family-level classifications between the number of pollen counted microscopically (grains/m³) and the BLAST analysis of the Illumina sequence reads for airborne pollen samples. Sequence read data has been square-root transformed for presentation purposes only. (B) Heat map of OTU abundance for the most prominent family-level classifications. Families assigned a (*) symbol signify that they were also counted by light microscopy. The hue and tone of colour reflects the abundance of sequence read data with darker tones denoting higher sequence reads.

short-term temporal changes in the distribution of pollen genera and aerobiological composition over an entire pollen season.

Targeted sequencing of the pollen aerobiome demonstrated that *rbcL* metabarcoding has the capacity to aid the classification of taxa to the genus level in most cases. However, it cannot always be relied upon to produce a definitive classification, but provides researchers with a number of options from which they can select a suitable candidate based on knowledge of the site and regional species distributions. Metabarcoding showed that there are prominent plant families including Euphorbiaceae, Amaranthaceae, Plantaginaceae, Moraceae and Cannabaceae that are not being identified and enumerated by light microscopy. Part of the reason for this is the selection of pollen counted by microscopy based on recognised allergenic importance coupled with the time involved in identification and counting, which limits the diversity captured by this time-consuming traditional approach.

The results of this study prompt consideration of other families which have pollen previously associated with respiratory diseases including Plantaginaceae, to be included in future pollen monitoring

campaigns in this region. Many of the same families were identified by the metabarcoding data and earlier pollen monitoring studies of Green et al. (2002), in which airborne pollen from Poaceae taxa predominated. However, the percentage distribution of the other taxa previously described being moderately higher by metabarcoding than observed by Green et al. In relation to peak flowering times, our data corresponded with the data of Green et al. for five of the most prominent families described including Poaceae, Cupressaceae, Casuarinaceae, Myrtaceae and Urticaceae. Interestingly, a difference in abundance observed in the pollen concentration data and the metabarcoding data for this study was seen in the Asteraceae family, whereby, the proportion of the pollen counted comprised 12% versus 0.6% for metabarcoding. Green et al. (2002) reported a comparable level of airborne Asteraceae pollen comprising 0.72% of total annual pollen, closely matching the metabarcoding figures. The possibility that this pollen is being misidentified by light microscopy seems unlikely given the distinct surface morphology of Asteraceae pollen. While the presence of chloroplast DNA in pollen has clearly been demonstrated, it

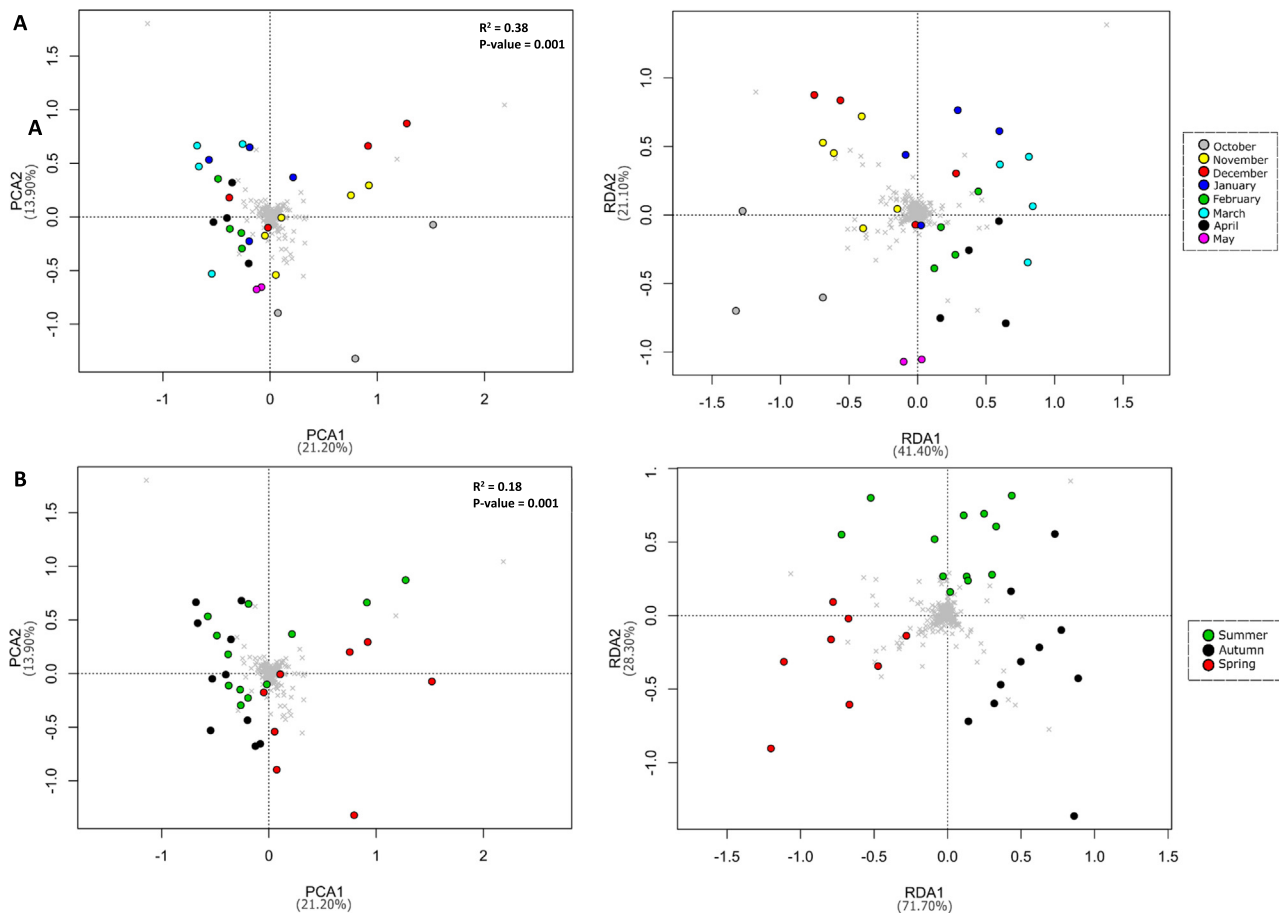


Fig. 5. Principal components analysis (PCA) and redundancy analysis (RDA) based on the relative abundance of OTUs during the 2016/17 Rocklea pollen season. (A) According to the calendar month and (B) According to the season.

remains unclear whether chloroplast DNA is present in the pollen of all seed plant species as it has been reported missing from the sperm cells in pollen from several taxa (Corriveau and Coleman, 1988; Zhang and Sodmergen, 2003; Bell et al., 2016b) including certain species of Asteraceae (Corriveau and Coleman, 1988). An alternative explanation may be that Asteraceae pollen is under-detected by metabarcoding due either to resistance to DNA extraction by the current protocol and/or lower amplification efficiency due to primer bias process (Elbrecht and Leese, 2015; Pornon et al., 2016; Pinol et al., 2018). Validating this observation, Bell et al. (2019) were frequently unable to detect common ragweed (*Ambrosia artemisiifolia* L.) with the *rbcl* marker gene. However, based on our in silico analysis of Asteraceae species typically found within the Brisbane region (<https://weeds.brisbane.qld.gov.au/search/weeds?keywords=Asteraceae>) it is clear that primer bias is unlikely to be a contributing factor, as the vast majority of *rbcl* sequences for both the forward and reverse primer should anneal (Fig. S4, Supporting information). These possibilities require further exploration given the allergenic importance of Asteraceae species including common ragweed (*Ambrosia artemisiifolia* L.) which is a notifiable weed in the Queensland region.

The capacity for DNA metabarcoding to be scaled up from a small experimental concept, to its application across a complete pollen season using routine equipment is valid and highly informative. However, the field is still struggling to quantify and appropriately communicate the level of confidence associated with individual plant taxonomic assignments based on metabarcoding. Various amplicons of the internal transcribed spacer (ITS) are considered as metabarcodes for plant species classification (Chen et al., 2010; Cheng et al., 2016). Brennan et al. (2019) observed that the ITS2 marker showed higher specificity

between species but did not detect all plants compared to the *rbcl* marker, which was highly universal, but the latter marker showed lower resolution between closely related plants. The inclusion of a second barcode such as *matK* was once considered necessary for accurate species level classification (de Vere et al., 2015), however, *matK* itself has a similar low substitution rate comparable to *rbcl* and single barcodes such as ITS2 among others can accurately match to species level (Kress et al., 2005; Li et al., 2011). Realistically, it is not cost effective to implement metabarcoding applications with two markers for routine airborne pollen monitoring across an entire season/s. This is because costs rapidly compound when analysing multiple samples in conjunction with the difficulty to associate the sequences from the two separate markers in a pooled sample to the same pollen grain. Because of this limitation, this study was confined to a single barcode for seasonal aerobiome analysis. Selection of a single marker for sequencing can lead to bioinformatic limitations, particularly with the merging of forward and reverse reads. Our experience, echoes Bell et al. (2016b), that the constraints imposed by the read length limit of illumina sequencing (300 bp), meaning the overlap between paired ends can be as little as 10 bp with which to assemble the reads.

Using relative pollen DNA sequence abundance alone as a response variable, it was possible to discriminate by calendar month and seasons while monitoring temporal change in the distribution of pollen genera and aerobiological composition. In keeping with the observations of Bell et al. (2019), who found statistically significant correlations for *rbcl* at the genus, species and families level ($p < .001$), when comparing the number of DNA sequencing reads for each species to the known proportions of the species, we also found such a relationship when examining 'month' and 'season' as factors. In spite of these being significant

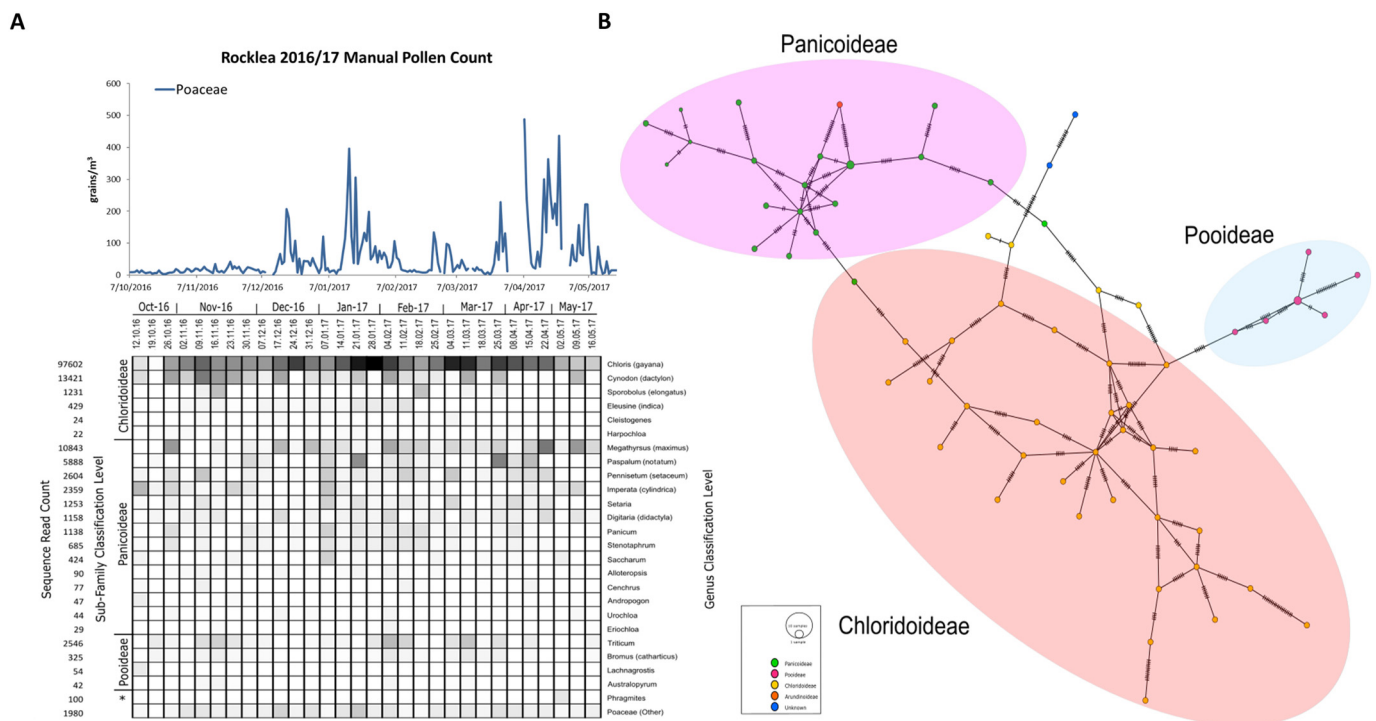


Fig. 6. (A) Heat map of OTU abundance for all grass genera from the Rocklea aerobiome, compared to pollen concentrations (grains/m³) counted by light microscopy for the Poaceae family. The hue and tone of colour reflects the abundance of sequence read data with darker tones denoting higher sequence reads. (B) Median-joining network of operational taxonomic units (OTU) present within the Poaceae family clustered utilising the *rbcL* metabarcode. Each individual node represents an individual OTU and each different colour denotes a different sub-family-group.

correlations, the R^2 for our study (0.38) and Bell et al. (2019) ($R^2 = 0.328$ and slope = 0.411) was weak for *rbcL*. Similarly, Brennan et al. (2019) recently reported temporal patterns in distributions of specific pollen by metabarcoding at temperate climate sites in Britain collected using a cyclonic sampling device and in their study, month of the year was a factor that improved their ability to predict taxonomic composition across the pollen season. While in this context, it is clear we extracted useful information using DNA sequence abundance, it must be stated that there are limitations in obtaining accurate quantitative information from read counts per se. Bell et al. (2016a) in their review of DNA barcoding methods in forensic palynology, explain how the copy number of chloroplast DNA in pollen can vary across species. This can mean that amplification success of species in a mixed sample may not reflect their pollen abundance. While NGS can give a semi-quantitative measure of copy number of the amplified region within the sample for each species, to accurately translate copy number to the proportion of pollen grains in the sample requires information on the copy number of the chloroplast DNA per pollen grain of each species. With regards to inter sample variability, Bell et al. (2019) in their quantitative assessment of pollen DNA metabarcoding using constructed species mixtures, found there was little variation in the proportion of reads for each species among replicates for either *rbcL* or ITS2.

A second key motivation for this project was to better understand the patterns of exposure to pollen of subtropical grasses compared with temperate grasses. Given differences in their allergen components, allergy patients show specific cellular and IgE antibody responses to Panicoideae, Chloridoideae and Pooideae pollen demonstrating the need for estimates of airborne grass pollen concentrations at genus level in order to provide a better understanding of the risk of allergic respiratory disease (Davies et al., 2012).

Metabarcoding appeared to be useful for tracking the temporal changes in pollen distribution for grass species, consistent with previous pollen concentration data for this same subtropical area, whereby the grass pollen season showed multiple peaks, with summer/early autumn

representing peak flowering times, regularly extended beyond spring (Green et al., 2002; Medek et al., 2016). The data showed that multiple peaks in airborne grass pollen observed were predominantly due to repeated flowering of the same Poaceae genera over the season with approximately 66% of grass genera displaying this pattern. This observation was best exemplified by heatmap data which showed that Chloridoideae species such as *Chloris gayana* and *Cynodon dactylon*, and Panicoideae species such as *Megathyrus maximus* and *Pennisetum setaceum* had multiple peaks of airborne grass pollen across the season. However, there were also examples of several genera flowering at distinct times such as *Saccharum* sp. (Panicoideae) and *Sporobolus elongatus* (Chloridoideae).

The preponderance of multiple flowering peaks at subtropical latitudinal sites is likely due to the predominance of pollen from subtropical grasses which take their flowering cues primarily from rainfall and temperature rather than photoperiod, as seen with the temperate grasses (Tarumoto et al., 2005).

The ability to accurately dissect the grass species composition of the Rocklea site via metabarcoding of routinely collected airborne pollen samples enables more precise monitoring and better understanding of seasonal AR burden and what grass pollens are present. The use of a single barcode such as *rbcL* limits the ability to obtain species level identification for all taxa, however, this level of resolution may not be required for aerobiological grass pollen distribution mapping and forecasting for the purposes of managing AR.

Many of the grass species identified in this study have been corroborated by previous research conducted by Moss (1967) who compiled a flowering calendar of possible hay fever plants in Brisbane. According to her studies, the most abundant species and highest in pollen yield were *Chloris gayana* and *Cynodon dactylon*, which were both from the sub-family Chloridoideae and were also the most predominant in our study. This high representation of *Chloris gayana* is due to its abundance in the grassland area in the immediate vicinity of the pollen sampling site. Moreover, many other grass species,

particularly from the sub-family Panicoideae, were identified as being abundant and corresponded with our lists e.g. *Imperata cylindrica* var. *major*, *Paspalum* spp., *digitaria didactyla* etc. It should also be noted, that the metabarcoding data did not detect several native species of grasses observed during the 1967 survey of Rocklea, which led us to speculate that there has been local species diversity loss in this region. According to in silico analysis, the majority of these native grasses have *rbcl* sequences that match the 3'-clamp of the metabarcode primers, suggesting they should have been detected if present in the aerobiome (Fig. S5, Supporting information). In fact, some of these native grasses were detected in the aerobiome, including *Sporobolus elongatus*, and *Imperata cylindrica*. The phenomenon of species loss has been observed Australia wide with the introduction of improved pasture lines from Africa (e.g. *Chloris gayana*) for the livestock industry, which generally outcompete their native counterparts.

Flowering times in the subtropical locale of Rocklea were defined by a series of discrete grass pollen peaks within the summer and autumn months consistent with repeated flowering of subtropical grasses from several genera. With the accrual of sequencing data from subsequent pollen seasons as well as the application of the technique to other sites in Australia, including subtropical and temperate climates, a more accurate understanding of Australian pollen aerobiology is anticipated. In particular, knowledge of grass pollen aerobiological diversity and phenology will assist with development of evidenced-based local pollen forecasts to assist with public health management of grass pollen-induced health threats including epidemic asthma (Devadas et al., 2018; Thien et al., 2018).

The application of metabarcoding to the study of palynology has now allowed a detailed analysis of the pollen species and distributions comprising the pollen aerobiome. Herein, we have demonstrated that our analysis pipeline for DNA metabarcoding of airborne pollen samples, routinely collected using a Burkard volumetric pollen and spore trap, can provide increased resolution of temporal changes in the composition of environmental air samples surveyed throughout an entire pollen season of an exemplary subtropical region. Such information is essential if we are to monitor alterations in the distributions and composition of airborne pollen of allergenic significance, and more broadly ecological trends in pollen biodiversity with climate change.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2020.141189>.

CRedit authorship contribution statement

B.C. Campbell: Conceptualization, Writing - original draft, Formal analysis. **J. Al Kouba:** Investigation. **V. Timbrell:** Investigation. **M. Noor:** Investigation. **K. Massel:** Formal analysis, Writing - review & editing. **E.K. Gilding:** Formal analysis, Writing - review & editing. **N. Angel:** Investigation, Writing - review & editing. **B. Kemish:** Formal analysis, Writing - review & editing. **P. Hugenholtz:** Resources, Software, Writing - review & editing. **I.D. Godwin:** Conceptualization, Writing - review & editing, Funding acquisition. **J.M. Davies:** Supervision, Conceptualization, Formal analysis, Writing - review & editing, Writing - original draft, Funding acquisition, Resources.

Declaration of competing interest

The authors declare outside the scope of this research, Davies leads the NHMRC AusPollen Partnership Project (GNT 1116107) with matching cash and in kind co-sponsorship from The Australasian Society for Clinical Immunology and Allergy, Asthma Australia, Bureau of Meteorology, Commonwealth Scientific and Industrial Research Organisation, Stallergenes Australia, Federal Office of Meteorology and Climatology MeteoSwiss, Switzerland. Davies is an investigator of current National Foundation for Medical Research Innovation grant with co-

sponsorship from Abionic Switzerland. She is an inventor on patents assigned to QUT. Her institute has received Honorarium payments and travel expenses for education sessions and conference presentations from Stallergenes Australia and Wymedical in the last five years.

Acknowledgements

We thank the Queensland Department of Science and Environment Air Quality Monitoring Team for their support and for granting site access to the Rocklea Air Quality Monitoring Site. We would like to thank all the valuable technical contributions from the ACE sequencing team and QUT Davies Allergy Research Group who have made this project possible.

Funding information

Funding for this project has been provided by the following funding bodies: Queensland University of Technology (QUT Catapult Grant – Davies); The University of Queensland (UQ Collaboration and Industry Engagement Fund – Davies & Godwin); Australian Research Council (ARC Discovery- Davies DP 170101630), Australian Centre for Ecogenomics UQ Strategic Funding.

References

- Anderson, M.J., 2005. PERMANOVA: A Fortran Computer Program for Permutational Multivariate Analysis of Variance. Department of Statistics, University of Auckland, Auckland, New Zealand.
- Bandelt, H., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036>.
- Beggs, P.J., Bennett, C.M., 2011. Climate change, aeroallergens, natural particulates, and human health in Australia: state of the science and policy. *Asia Pac. J. Public Health* 23, 46–53. <https://doi.org/10.1177/1010539510391771>.
- Beggs, P. J., Davies, J.M., Milic, A., Haberle, S.G., Johnston, F.H., Jones, P.J., ... Newbigin, E. (2018). Australian airborne pollen and spore monitoring network interim standard and protocols. https://www.allergy.org.au/images/stories/pospapers/Australian_Pollen_and_Spore_Monitoring_Interim_Standard_and_Protocols_v2_14092018.pdf. Macquarie University and Queensland University of Technology. Accessed 03/01/2019.
- Bell, K.L., Burgess, K.S., Okamoto, K.C., Aranda, R., Brosi, B.J., 2016a. Review and future prospects for DNA barcoding methods in forensic palynology. *Forensic Sc. Int. Genet.* 21, 110–116. <https://doi.org/10.1016/j.fsigen.2015.12.010>.
- Bell, K.L., de Vere, N., Keller, A., Richardson, R.T., Gous, A., Burgess, K.S., Brosi, B.J., 2016b. Pollen DNA barcoding: current applications and future prospects. *Genome* 59, 629–640. <https://doi.org/10.1139/gen-2015-0200>.
- Bell, K.L., Burgess, K.S., Botsch, J.C., Dobbs, E.K., Read, T.D., Brosi, B.J., 2019. Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Mol. Ecol.* 28, 431–455. <https://doi.org/10.1111/mec.14840>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brennan, G.L., Potter, C., de Vere, N., Griffith, G.W., Skjoth, C.A., Osborne, N.J., ... Creer, S., 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution* 3, 750–754. <https://doi.org/10.1038/s41559-019-0849-7>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinform.* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>.
- Campbell, B.C., Gilding, E.K., Timbrell, V., Guru, P., Loo, D., Zennaro, D., ... Davies, J.M., 2015. Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. *Journal of Allergy and Clinical Immunology* 135, 133–142. <https://doi.org/10.1016/j.jaci.2014.06.034>.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. <https://doi.org/10.1038/nmeth.f303>.
- CBOL Plant Working Group A DNA barcode for land plants, 2009. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12794–12797. <https://doi.org/10.1073/pnas.0905845106>.
- Chabre, H., Gouyon, B., Huet, A., Baron-Bodo, V., Nony, E., Hrabina, M., ... Moingen, P., 2010. Molecular variability of group 1 and 5 grass pollen allergens between Poideae species: implications for immunotherapy. *Clinical and Experimental Allergy* 40, 505–519. <https://doi.org/10.1111/j.1365-2222.2009.03380.x>.
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., ... Leon, C., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5 (1), e8613. <https://doi.org/10.1371/journal.pone.0008613>.
- Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., Zhou, S., 2016. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Mol. Ecol. Resour.* 16, 138–149. <https://doi.org/10.1111/1755-0998.12438>.

- Corriveau, J.L., Coleman, A.W., 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am. J. Bot.* 75, 1443–1458. <https://doi.org/10.2307/2444695>.
- Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* 29, 566–571. <https://doi.org/10.1016/j.tree.2014.08.001>.
- Davies, J.M., Bright, M.L., Rolland, J.M., O'Hehir, R.E., 2005. Bahia grass pollen specific IgE is common in seasonal rhinitis patients but has limited cross-reactivity with ryegrass. *Allergy* 60, 251–255. <https://doi.org/10.1111/j.1398-9995.2005.00663.x>.
- Davies, J.M., Mittag, D., Dang, T.D., Symons, K., Voskamp, A., Rolland, J.M., O'Hehir, R.E., 2008. Molecular cloning, expression and immunological characterisation of Pas n 1, the major allergen of Bahia grass *Paspalum notatum* pollen. *Mol. Immunol.* 46, 286–293. <https://doi.org/10.1016/j.molimm.2008.08.267>.
- Davies, J.M., Dang, T.D., Voskamp, A., Drew, A.C., Biondo, M., Phung, M., O'Hehir, R.E., 2011a. Functional immunoglobulin E cross-reactivity between Pas n 1 of Bahia grass pollen and other group 1 grass pollen allergens. *Clin. Exp. Allergy* 41, 281–291. <https://doi.org/10.1111/j.1365-2222.2010.03670.x>.
- Davies, J.M., Voskamp, A., Dang, T.D., Pettit, B., Loo, D., Petersen, A., O'Hehir, R.E., 2011b. The dominant 55 kDa allergen of the subtropical Bahia grass (*Paspalum notatum*) pollen is a group 13 pollen allergen, Pas n 13. *Mol. Immunol.* 48, 931–940. <https://doi.org/10.1016/j.molimm.2010.12.013>.
- Davies, J.M., Li, H., Green, M., Towers, M., Upham, J.W., 2012. Subtropical grass pollen allergens are important for allergic respiratory diseases in subtropical regions. *Clin. Trans. Allergy* 2 (4). <https://doi.org/10.1186/2045-7022-4>.
- de Vere, N., Rich, T.C., Trinder, S.A., Long, C., 2015. DNA barcoding for plants. *Methods Mol. Biol.* 1245, 101–118. https://doi.org/10.1007/978-1-4939-1966-6_8.
- De Weger, L.A., Beerthuis, T., Gast-Strookman, J.M., van der Plas, D.T., Terreehorst, I., Hiemstra, P.S., Sont, J.K., 2011. Difference in symptom severity between early and late grass pollen season in patients with seasonal allergic rhinitis. *Clin. Trans. Allergy* 1, 18. <https://doi.org/10.1186/2045-7022-18>.
- Devadas, R., Huete, A.R., Vicendese, D., Erbas, B., Beggs, P.J., Medek, D., Davies, J.M., 2018. Dynamic ecological observations from satellites inform aerobiology of allergenic grass pollen. *Sci. Total Environ.* 633, 441–451. <https://doi.org/10.1016/j.scitotenv.2018.03.191>.
- Elbrecht, V., Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass–sequence relationships with an innovative metabarcoding protocol. *PLoS One* 10, e0130324. <https://doi.org/10.1371/journal.pone.0130324>.
- Etto, T., De Boer, C., Prickett, S., Gardner, L.M., Voskamp, A., Davies, J.M., Rolland, J.M., 2012. Unique and cross-reactive T cell epitope peptides of the major Bahia grass pollen allergen, pas n 1. *Int. Arch. Allergy Immunol.* 159, 355–366. <https://doi.org/10.1159/000338290>.
- Green, B.J., Dettmann, M.E., Rutherford, S., Simpson, R.W., 2002. Airborne pollen of Brisbane, Australia: a five-year record, 1994–1999. *Grana* 41, 242–250. <https://doi.org/10.1080/001731302321012031>.
- Hirst, J.M., 1952. An automatic volumetric spore trap. *Ann. Appl. Biol.* 39, 257–265. <https://doi.org/10.1111/j.1744-7348.1952.tb00904.x>.
- Kailaivasan, T.H., Timbrell, V.L., Solley, G., Smith, W.B., McLean-Tooke, A., van Nunen, S., ... Davies, J.M. (2020). Biogeographical variation in specific IgE recognition of temperate and subtropical grass pollen allergens in allergic rhinitis patients. *Clin. Trans. Immunol.* 9(2):e01103. doi: <https://doi.org/10.1002/cti2.1103>.
- Katelaris, C.H., Lee, B.W., Potter, P.C., Maspero, J.F., Cingi, C., Lopatin, A., Walters, R.D., 2012. Prevalence and diversity of allergic rhinitis in regions of the world beyond Europe and North America. *Clin. Exp. Allergy* 42, 186–207. <https://doi.org/10.1111/j.1365-2222.2011.03891.x>.
- Keller, A., Danner, N., Grimmer, G., Ankenbrand, M., von der Ohe, K., von der Ohe, W., Steffan-Dewenter, I., 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. *Plant Biol.* 17, 558–566. <https://doi.org/10.1111/plb.12251>.
- Kraaijeveld, K., de Weger, L.A., Ventayol Garc'a, M., Buermans, H., Frank, J., Hiemstra, P.S. & den Dunnen, J.T. (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Mol. Ecol. Resour.*, 15, 8–16. doi: <https://doi.org/10.1111/1755-0998.12288>.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374.
- Landsmeer, S.H., Hendriks, E.A., de Weger, L.A., Reiber, J.H.C., Stoel, B.C., 2009. Detection of pollen grains in multifocal optical microscopy images of air samples. *Microsc. Res. Tech.* 72, 424–430. <https://doi.org/10.1002/jemt.20688>.
- Leigh, J.W., Bryant, D., 2015. PopART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. <https://doi.org/10.1111/2041-210X.12410>.
- Leontidou, K., Vernesi, C., De Groeve, J., Cristofolini, F., Vokou, D., Cristofori, A., 2018. DNA metabarcoding of airborne pollen: new protocols for improved taxonomic identification of environmental samples. *Aerobiologia* 34, 63–74. <https://doi.org/10.1007/s10453-017-9497-z>.
- Li, D.Z., Gao, L.M., Li, H.T., Wang, H., Ge, X.J., Liu, J.Q., ... Duan, G.W. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. U. S. A.*, 108, 19641–19646. doi: <https://doi.org/10.1073/pnas.1104551108>.
- Longhi, S., Cristofori, A., Gatto, P., Cristofolini, F., Grando, M.S., Gottardini, E., 2009. Biomolecular identification of allergenic pollen: a new perspective for aerobiological monitoring? *Ann. Allergy Asthma Immunol.* 103, 508–514. [https://doi.org/10.1016/S1081-1206\(10\)60268-2](https://doi.org/10.1016/S1081-1206(10)60268-2).
- Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G., Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinform.* 13, 31. <https://doi.org/10.1186/1471-2105-13-31>.
- Medek, D.E., Beggs, P.J., Erbas, B., Jaggard, A.K., Campbell, B., Vicendese, D., Davies, J.M., 2016. Regional and seasonal variation in airborne grass pollen levels between cities of Australia and New Zealand. *Aerobiologia* 32, 289–302. <https://doi.org/10.1007/s10453-015-9399-x>.
- Morgan, J.A., LeCain, D.R., Pendall, E., Blumenthal, D.M., Kimball, B.A., Carrillo, Y., West, M., 2011. C4 grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland. *Nature* 476, 202–205. <https://doi.org/10.1038/nature10274>.
- Moss, J.E., 1967. A flowering calendar of possible hay fever plants in Brisbane. *Med. J. Aust.* 1, 270–272.
- Nony, E., Timbrell, V., Hrabina, M., Boutron, M., Solley, G., Moingeon, P., Davies, J.M., 2015. Specific IgE recognition of pollen allergens from subtropical grasses in patients from the subtropics. *Ann. Allergy Asthma Immunol.* 114, 214–220. <https://doi.org/10.1016/j.jana.2014.12.005>.
- Núñez, A., Amo de Paz, G., Ferencova, Z., Rastrojo, A., Guantes, R., García, A.M., Alcami, A., ... Moreno, D.A. (2017). Validation of the first-type spore trap for simultaneous monitoring of prokaryotic and eukaryotic biodiversities in urban air samples by next-generation sequencing. *Appl. Environ. Microbiol.*, 83, e00472–17. doi: <https://doi.org/10.1128/AEM.00472-17>.
- Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P.L., Minchin, P.R., Hara, R.B., ... Wagner, H., 2013. Vegan: community ecology package. R package version 2.0-10 [online]. URL: <http://CRAN.R-project.org/package=vegan>.
- Pawankar, R., Canonica, G.W., Holgate, S., Lockey, R.F., 2011. *World Allergy Organization White Book on Allergy*. World Allergy Organization, Milwaukee (WI).
- Pinol, J., Senar, M.A., Symondson, W., 2018. The choice of universal primers and the characteristics of the species mixture determines when DNA metabarcoding can be quantitative. *Mol. Ecol.* 28, 407–419. <https://doi.org/10.1111/mec.14776>.
- Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Andaló, C., 2016. Using metabarcoding to reveal and quantify plant-pollinator interactions. *Sci. Rep.* 6, 27282. <https://doi.org/10.1038/srep27282>.
- Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5 (3), e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- Rideout, J.R., He, Y., Navas-Molina, J.A., Walters, W.A., Ursell, L.K., Gibbons, S.M., ... Caporaso, J.G. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ*, 21, 2:e545. doi: <https://doi.org/10.7717/peerj.545>.
- Rittenour, W.R., Hamilton, R.G., Beezhold, D.H., Green, B.J., 2012. Immunologic, spectrophotometric and nucleic acid based methods for the detection and quantification of airborne pollen. *J. Immunol. Methods* 383, 47–53. <https://doi.org/10.1016/j.jim.2012.01.012>.
- Roberts, D.W., 2012. Package 'labdsv'. URL: <https://CRAN.R-project.org/package=labdsv>.
- Rogers, S.O., Bendich, A.J., 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant-tissues. *Plant Mol. Biol.* 5, 69–76. <https://doi.org/10.1007/BF00020088>.
- Tang, L.Y., Nagata, N., Matsushima, R., Chen, Y., Yoshioka, Y., Sakamoto, W., 2009. Visualization of plastids in pollen grains: involvement of FtsZ1 in pollen plastid division. *Plant Cell Physiol.* 50, 904–908. <https://doi.org/10.1093/pcp/pcp042>.
- Tarumoto, I., Sato, H., Horibata, T., 2005. The effect of temperature and daylength on heading in tetraploid cultivars of Rhodesgrass (*Chloris gayana* Kunth). *Grassl. Sci.* 51, 281–285. <https://doi.org/10.1111/j.1744-697X.2005.00037.x>.
- Thibaudon, M., Caillaud, D., Besancenot, J.P., 2013. Methods of studying airborne pollen and pollen calendars. *La Revue des Maladies Respiratoires* 30, 463–479. <https://doi.org/10.1016/j.rmr.2013.02.006>.
- Thien, F., Beggs, P.J., Csutoros, D., Darvall, J., Hew, M., Davies, J.M., Guest, C., 2018. The Melbourne epidemic thunderstorm asthma event 2016: an investigation of environmental triggers, effect on health services, and patient risk factors. *Lancet. Planet. Health* 2, e255–e263. [https://doi.org/10.1016/S2542-5196\(18\)30120-7](https://doi.org/10.1016/S2542-5196(18)30120-7).
- Weber, R.W., 2007. Cross-reactivity of pollen allergens: impact on allergen immunotherapy. *Ann. Allergy Asthma Immunol.* 99, 203–211. [https://doi.org/10.1016/S1081-1206\(10\)60654-0](https://doi.org/10.1016/S1081-1206(10)60654-0).
- Zhang, Q., Liu, Y., Sodmergen, 2003. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant Cell Physiol.* 44, 941–951. <https://doi.org/10.1093/pcp/pcg121>.

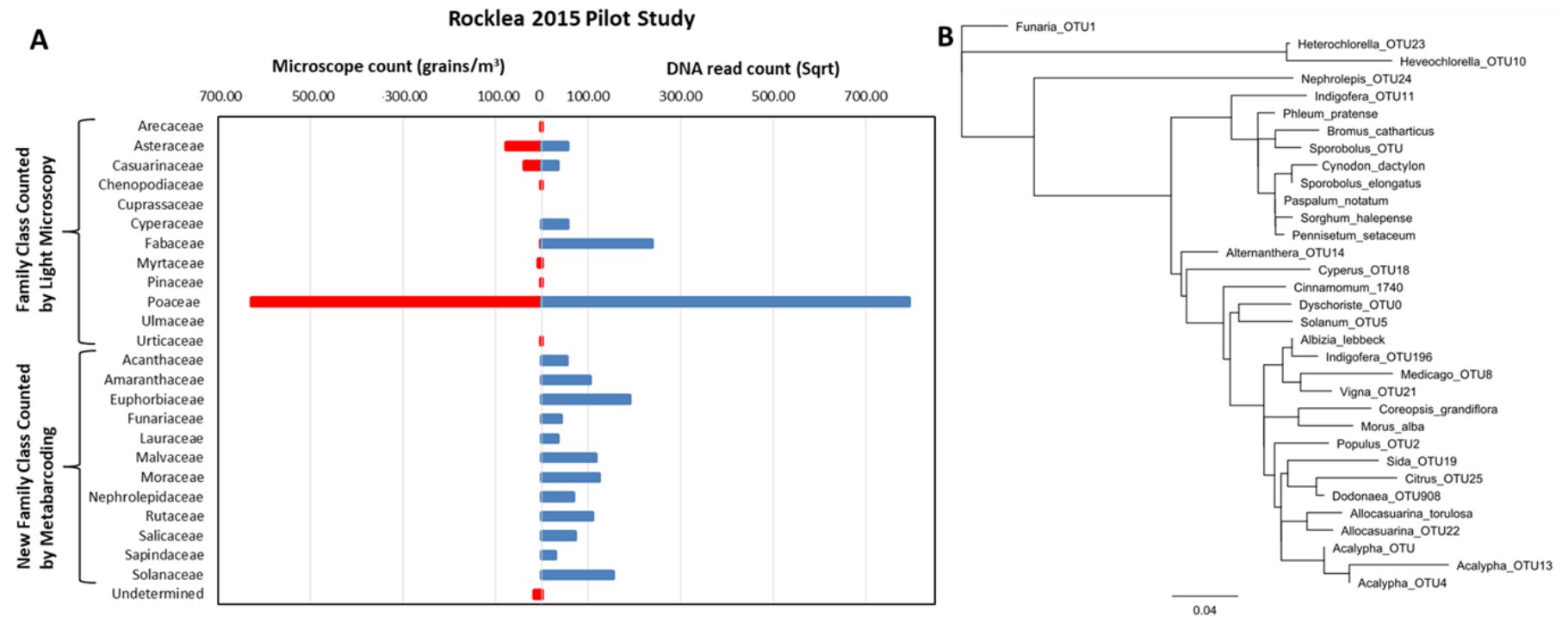


Fig. S1, Supp. Information 2015 Rocklea pilot study data **(A)** Comparison of family-level classifications between the number of pollen counted microscopically (grains/m³) and the BLAST analysis of the Illumina sequence reads for airborne pollen samples. Sequence read data has been square-root transformed for presentation purposes only. **(B)** Tree diagram for cluster analysis of individual OTU classified down to the genus-level from 12 environmental samples and 1 control. Note: *Phleum pratense*; *Sorghum halepense*; and *Cynodon dactylon* species were added as the mixed grass control.

Tracking Seasonal Changes in Diversity of Pollen Allergen Exposure: Targeted Metabarcoding of a Subtropical Aerobiome

Campbell, B.C., Al Kouba, J., Timbrell, V., Noor, M., Massel, K., Gilding, E.K., Angel, N., Kemish, B., Hugenholz, P., Godwin, I.D. and Davies, J.M.

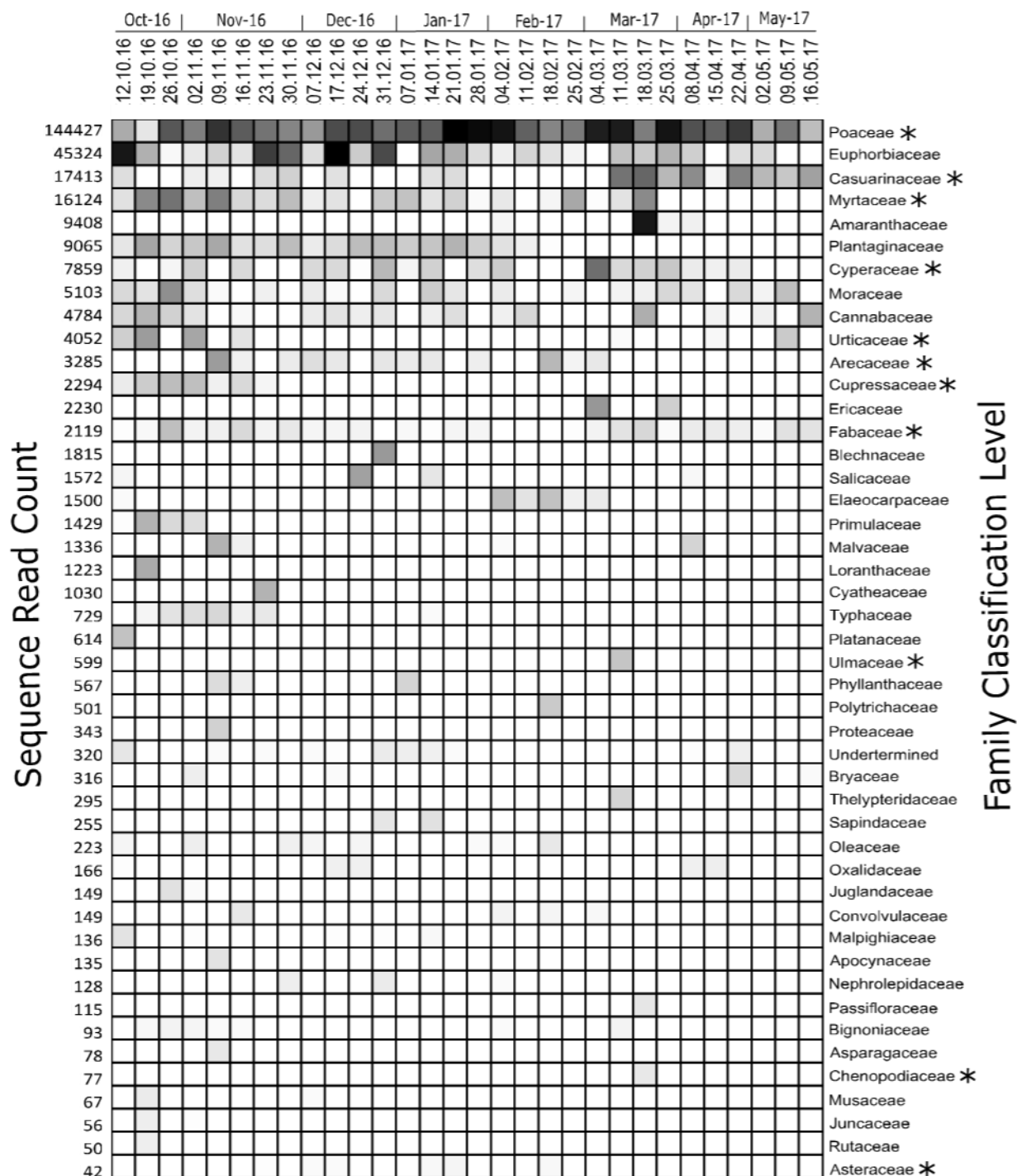


Fig. S2, Supp. Information Heat map of OTU abundance for family-level classifications for the 2016/17 Rocklea aerobiome. Genera hits for $\approx \leq 50$ reads have not been displayed. Families assigned a (*) symbol signify that they were also counted by light microscopy. The hue and tone of colour reflects the abundance of sequence read data with darker tones denoting higher sequence reads.

Tracking Seasonal Changes in Diversity of Pollen Allergen Exposure: Targeted Metabarcoding of a Subtropical Aerobiome
 Campbell, B.C., Al Kouba, J., Timbrell, V., Noor, M., Massel, K., Gilding, E.K., Angel, N., Kemish, B., Hugenholz, P., Godwin, I.D. and Davies, J.M.

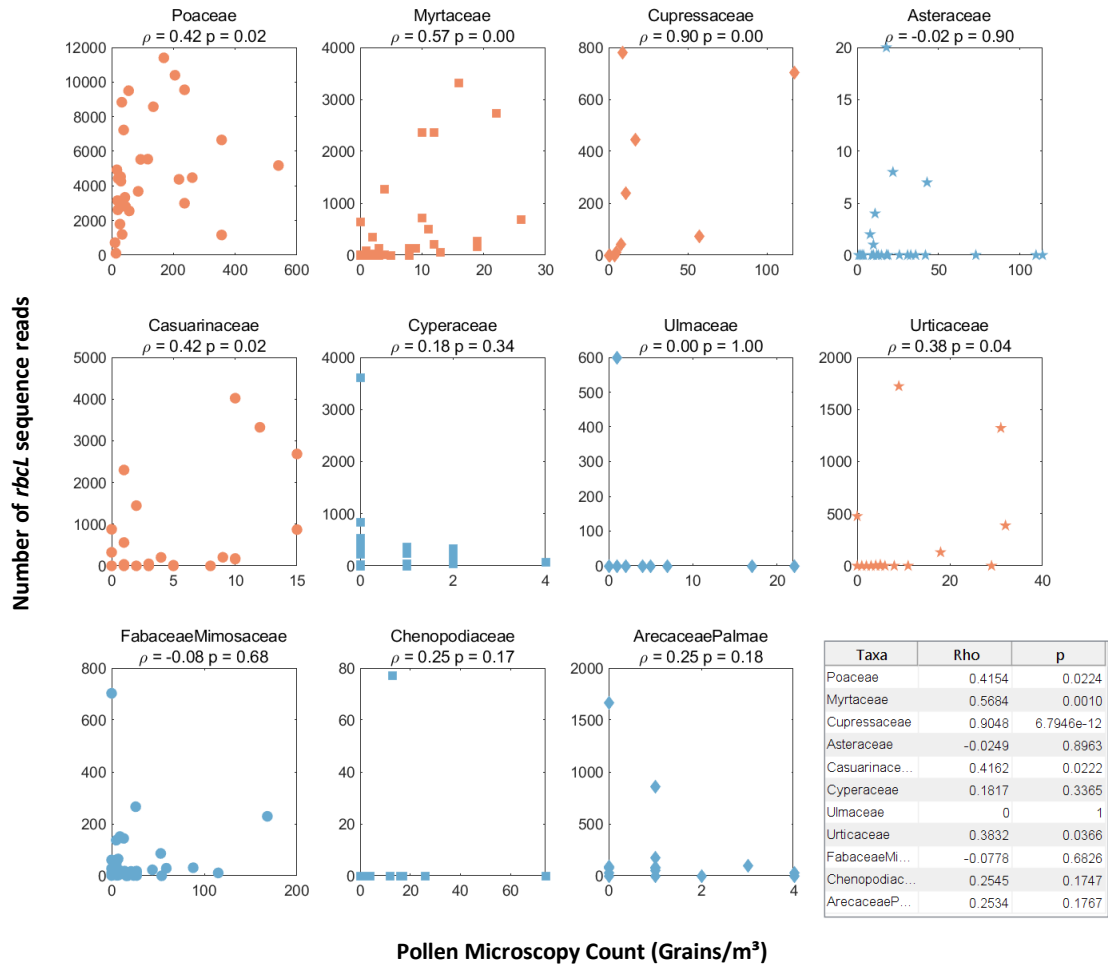


Fig. S3, Supp. Information Spearman's rank correlation coefficient between the numbers of *rbcL* sequence reads and pollen concentrations (Grains/m³) according to prominent families counted microscopically. A value of $p < 0.05$ was considered statistically significant and values coloured orange.

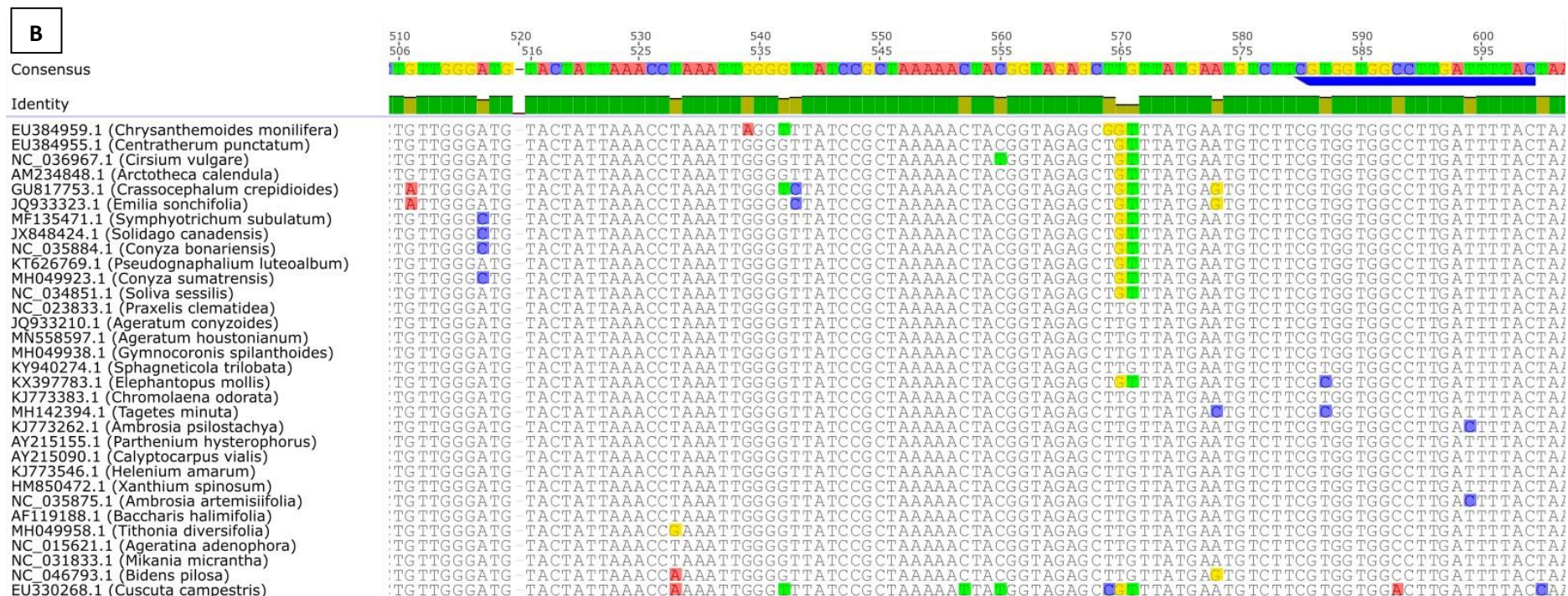


Fig. S4, Supp. Information *In silico* analysis showing lack of primer bias of Asteraceae species of the CBOL (2009) *rbcl* DNA barcode, believed to exist in the greater Rocklea region as listed by the Brisbane City Council Weed Identification website (<https://weeds.brisbane.qld.gov.au/search/weeds?keywords=Asteraceae>) using *rbcl* sequences sourced from NCBI as of the 03/06/2020. (A) Multiple sequence alignment using ClustalW against the forward metabarcoding primer for *rbcl* annotated with a blue arrow. (B) Multiple sequence alignment using ClustalW against the reverse metabarcoding primer for *rbcl* annotated with a blue arrow. All SNPs are highlighted in sequences.

A

	1	10	20	30	40	50	60	70	80
	1	10	20	30	39	49	59	69	79
Consensus	A	G	C	A	C	C	A	C	A
Identity	A	G	C	A	C	C	A	C	A
KM538809.1 (Capillipedium parviflorum)	A	T	G	T	C	A	C	C	A
NC_040131.1 (Bothriochloa decipiens)	A	T	G	T	C	A	C	C	A
LS398103.1 (Sorghum arundinaceum)	A	T	G	T	C	A	C	C	A
NC_035027.1 (Heteropogon contortus)	A	T	G	T	C	A	C	C	A
JN407343.1 (Imperata cylindrica)	A	T	G	T	C	A	C	C	A
KP827659.1 (Zea mays)	A	T	G	T	C	A	C	C	A
HQ600391.1 (Paspalum dilatatum)	A	T	G	T	C	A	C	C	A
EF125136.1 (Rhynchelytrum repens)	A	T	G	T	C	A	C	C	A
LN907967.1 (Megathyrsus maximus)	A	T	G	T	C	A	C	C	A
FN870409.1 (Setaria parviflora)	A	T	G	T	C	A	C	C	A
GU135184.1 (Pennisetum setaceum)	A	T	G	T	C	A	C	C	A
MK188834.1 (Megathyrsus maximus)	A	T	G	T	C	A	C	C	A
HQ600364.1 (Sporobolus elongatus)	A	T	G	T	C	A	C	C	A
MF114355.1 (Cynodon dactylon)	A	T	G	T	C	A	C	C	A
MF998383.1 (Chloris gayana)	A	T	G	T	C	A	C	C	A
NC_030486.1 (Eleusine indica)	A	T	G	T	C	A	C	C	A
HQ600436.1 (Bromus catharticus)	A	T	G	T	C	A	C	C	A
KX290938.1 (Themeda triandra)	A	T	G	T	C	A	C	C	A
NC_036973.1 (Poa annua)	A	T	G	T	C	A	C	C	A

B

Consensus

Identity

KM538809.1 (Capillipedium parviflorum)
 NC_040131.1 (Bothriochloa decipiens)
 LS398103.1 (Sorghum arundinaceum)
 NC_035027.1 (Heteropogon contortus)
 JN407343.1 (Imperata cylindrica)
 KP827659.1 (Zea mays)
 HQ600391.1 (Paspalum dilatatum)
 KJ773726.1 (Paspalum urvillei)
 HE573427.1 (Cenchrus setaceus)
 AM849404.1 (Digitaria didactyla)
 EF125136.1 (Rhynchelytrum repens)
 LN907967.1 (Megathyrsus maximus)
 FN870409.1 (Setaria parviflora)
 GU135184.1 (Pennisetum setaceum)
 HQ600364.1 (Sporobolus elongatus)
 MF114355.1 (Cynodon dactylon)
 MF998383.1 (Chloris gayana)
 LN907943.1 (Eragrostis tenuifolia)
 NC_030486.1 (Eleusine indica)
 HQ600436.1 (Bromus catharticus)
 KX290938.1 (Themeda triandra)
 NC_036973.1 (Poa annua)

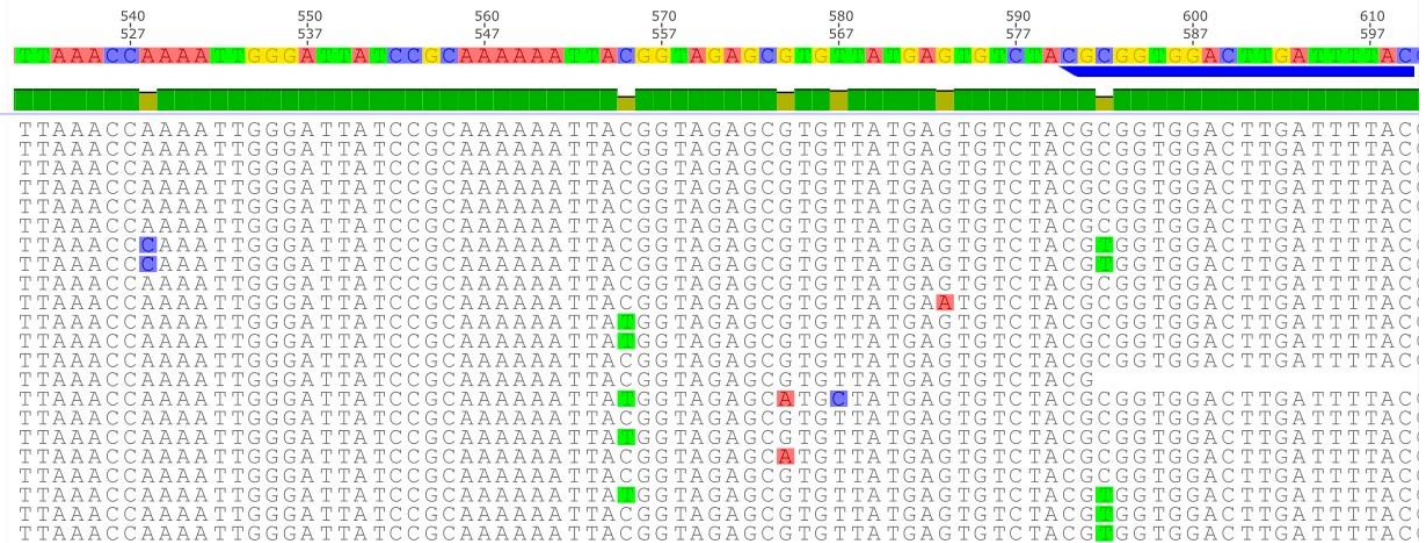


Fig. S5, Supp. Information *In silico* analysis showing lack of primer bias of Australian grasses, both native and introduced, of the CBOL (2009) *rbcl* DNA barcode as identified by Moss (1967)* in the Rocklea study region using *rbcl* sequences sourced from NCBI as of the 03/06/2020. (A) Multiple sequence alignment using ClustalW against the forward metabarcoding primer for *rbcl* annotated with a blue arrow. (B) Multiple sequence alignment using ClustalW against the reverse metabarcoding primer for *rbcl* annotated with a blue arrow. All SNPs are highlighted in sequences. *Moss, J.E. (1967). A flowering calendar of possible hay fever plants in Brisbane. The Medical Journal of Australia, 1, 270-272.